

基于拉普拉斯算子的人脸表情克隆优化方法

张满囤¹, 葛新杰¹, 霍江雷¹, 肖智东², 游理华², 张建军²

ZHANG Mandun¹, GE Xinjie¹, HUO Jianglei¹, XIAO Zhidong², YOU Lihua², ZHANG Jianjun²

1.河北工业大学 计算机科学与软件学院, 天津 300401

2.伯恩茅斯大学 英国国家动画中心, 英国 伯恩茅斯 BH12 5BB

1.School of Computer Science and Engineering, Hebei University of Technology, Tianjin 300401, China

2.National Centre for Computer Animation, Bournemouth University, Bournemouth BH12 5BB, UK

ZHANG Mandun, GE Xinjie, HUO Jianglei, et al. Facial expression cloning optimization method based on Laplace operator. *Computer Engineering and Applications*, 2016, 52(6): 178-181.

Abstract: In view of the reality of facial expression cloning and efficiency of expression reconstruction, a novel method based on motion capture data is proposed. After capturing the data of six fundamental expressions, it normalizes these data to make them in the same range. Then 41 points are chosen in critical areas of facial expression and it gets cloning expression using Laplace deformation algorithm with convex weight which can preserve the details of facial expression to avoid the low fidelity of uniform weights and unstable calculation of cotangent weights. Experimental results show that this method can generate realistic and natural expression animations and the efficiency of facial expression cloning is improved significantly.

Key words: motion capture; expression cloning; Laplace operator; facial expression

摘要:针对人脸克隆表情的真实性和表情重建的效率,提出了一种基于运动捕捉数据的人脸表情克隆新方法。使用运动捕捉设备捕捉人脸六种基本表情数据,并对数据进行归一化处理,使其位于同一区间范围,保证克隆表情数据的有效性;在表现人脸表情的关键区域选取41个特征点,通过采用凸权值的拉普拉斯变形算法将人脸表情转移到目标模型上,很好保留了人脸表情的细节,克服了均一权值保真度低和余切权值计算不稳定的问题。实验结果表明,该方法合成的克隆表情真实、自然,人脸表情克隆的效率得到了显著提高。

关键词:运动捕捉;表情克隆;拉普拉斯算子;人脸表情

文献标志码:A **中图分类号:**TP391.41 **doi:**10.3778/j.issn.1002-8331.1403-0338

1 引言

真实感的人脸表情克隆^[1-2]技术是近年来计算机动画、计算机游戏、计算机视觉领域的研究热点,并广泛应用于影视特效的制作,例如国外推出的电影《冰河世纪》、《蜘蛛侠》等。基于数据驱动的人脸表情克隆技术充分利用已有的顶点运动捕捉数据可以快速合成克隆表情,但是克隆表情后期经常需要对眼睛、牙齿这些脸

部凹陷的区域作后期处理,而且变形过程中往往需要融合多种方法,制作要求较高。从Zhang和Liu^[3-4]的工作以来,许多研究者们投入到面向普通用户的人脸表情克隆技术研究中,而且已经取得了显著进展。Noh和Neumann^[5]利用已有顶点的动画数据,提出了一种基于径向基函数^[6-7](RBF)的表情克隆方法,这种方法仅仅转移了运动矢量,而且对于稀疏的源数据,将会丢失源模型的

基金项目:河北省引进留学人员资助项目(No.C2013003015)。

作者简介:张满囤(1971—),男,博士,副教授,主要研究方向为人机交互、计算机图形学, E-mail: zhangmandun@scse.hebut.edu.cn;

葛新杰(1987—),男,硕士研究生,主要研究方向为人脸动画、图形图像处理;霍江雷(1989—),男,硕士研究生,主要研究

方向为计算机动画,计算机图形学;肖智东(1972—),男,博士,讲师,主要研究方向为运动捕捉及计算机动画;游理华

(1954—),男,博士,高级讲师,主要研究方向为计算机建模和仿真;张建军(1959—),男,博士,教授,主要研究方向为三维虚拟人动画、可视化和基于物理的仿真。

收稿日期:2014-03-24 **修回日期:**2014-06-04 **文章编号:**1002-8331(2016)06-0178-04

CNKI网络优先出版:2014-08-15, <http://www.cnki.net/kcms/doi/10.3778/j.issn.1002-8331.1403-0338.html>

部分表情细节;Deng 和 Neumann^[8]通过预先记录的人脸运动捕捉数据库,提出一种基于特征点的人脸表情动画方法来合成真实的表情动画,然而动画的质量严重依赖于庞大的数据库;Seol 和 Lewis 等人^[9]通过增加边界约束条件,将重定义问题用泊松方程来表示,利用贝叶斯公式产生与源运动一致的目标动画,这种方法仅集中于大规模的脸部变形,没有考虑现实的皱纹这些小尺度细节;Bhat 和 Goldenthal 等人^[10]提出了一种基于人脸重要轮廓特征的表演捕捉方法,该方法主要是对眼睛、嘴部这些人脸表情的关键区域进行表情匹配处理来得到高质量的克隆表情,这种方法对运动捕捉设备有较高的要求;Wan 和 Jin^[11]提出了一种基于拉普拉斯^[12-13]变形的数据驱动的人脸表情合成方法,对于顶点数目较多的模型,该方法的处理速度较慢,同时没有考虑复杂的眼睛和牙齿运动。本文在分析和总结前人工作的基础上,对拉普拉斯变形算法进行了改进。

本文采用了一种凸权值代替均一权值^[13]和余切权值^[14]的拉普拉斯变形算法,该算法易于实现,能够解决径向基函数插值算法出现奇异情况的问题,避免了表情重建后对眼睛、牙齿的细节处理,有效提高了表情克隆的真实性和生成效率。

2 基于凸权值的拉普拉斯变形算法

对每个 $i=1, 2, \dots, n$, 定义 $N(i)$ 为顶点 v_i 所有邻接点的集合, d_i 为集合 $N(i)$ 中元素的个数, 网格顶点的微分坐标^[13]可以表示为该顶点与其邻接顶点坐标加权求和的差:

$$\delta_i = (\delta_i^x, \delta_i^y, \delta_i^z) = v_i - \frac{1}{d_i} \sum_{j \in N(i)} v_j \quad (1)$$

记 I 为单位矩阵, A 为网格的邻接矩阵, D 为对角矩阵, 且 $D_{ii} = d_i$, 用矩阵表示式(1):

$$L = I - D^{-1}A \quad (2)$$

如图 1, Meyer 等人^[14]提出采用余切权值代替式(1)的均一权值:

$$\delta_i^c = \frac{1}{|\Omega_i|} \sum_{j \in N(i)} \frac{1}{2} (\cot \alpha_{ij} + \cot \beta_{ij}) (v_i - v_j) \quad (3)$$

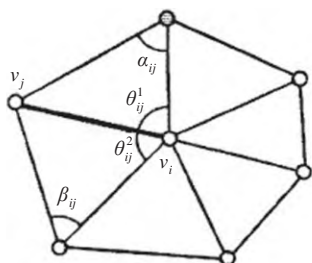


图1 余切权值中角的定义及边 (i, j) 的平均值坐标

其中, α_{ij}, β_{ij} 为边 (i, j) 所对应的两个角, $|\Omega_i|$ 为顶点 v_i 处 Voronoi 晶格^[14]的大小。

δ_i 只与网格的拓扑相关, 而 δ_i^c 依赖于网格的几何结构。但余切权值中的余切值可能为负, 并且当角度接近于 180° 时计算会不稳定, 为此, 本文采用模仿余切权值的凸权值来计算每个顶点的贡献:

$$w_{ij} = \frac{\tan(\theta_{ij}^1/2) + \tan(\theta_{ij}^2/2)}{\|v_i - v_j\|} \quad (4)$$

最后, 通过隐式求解的方法^[13]将微分坐标进行变换, 使得网格的变形与重建同时得到实现。采用最小二乘法, 可以求解所有顶点的坐标:

$$\tilde{V} = \arg \min_v (\|LV - \delta\|^2 + \sum_{j=1}^m w_j^2 \|v_j - c_j\|^2) \quad (5)$$

其中, \tilde{V} 为目标模型表情克隆后的所有顶点坐标, V 为初始坐标, δ 为 V 对应的微分坐标, w_j 为顶点 v_j 的权重, c_j 为目标模型上第 j 个特征点。

3 数据驱动的人脸克隆表情生成

3.1 运动数据的捕捉和处理

本文采集运动捕捉数据所用的是 Motion Analysis 公司的运动捕捉设备。按照 MPEG-4^[15]标准, 在人脸关键表情区域放置 41 个标记点, 捕捉表演者六种基本表情: 生气、悲伤、厌恶、吃惊、恐惧和高兴, 这些表情数据带有 41 个特征点的 3D 坐标关键帧信息。由于在一个非常有限的面部区域捕捉稠密的点云, 运动捕捉标记不可避免地会出现丢失, 因此需要进行后期数据处理, 包括原始数据标记标号、数据清理、间隙填充和数据归一化。

为了确定标记能够被捕捉设备识别, 需要在捕捉六种基本表情之前预先捕捉一系列面部运动序列, 它将为运动捕捉系统确定每一个标记的可能位置提供足够的信息。为了进行有效的数据清理, 需要通过标记的布局建立一组关系, 在这些关系中, 确定的标记被看作静止的参考点, 例如前额、鼻尖、下巴等。对于丢失的标记, 采用 4 个标记作为一组来估计和确定它们在数据间隙中的位置。图 2 所示是通过运动捕捉设备重建的人脸



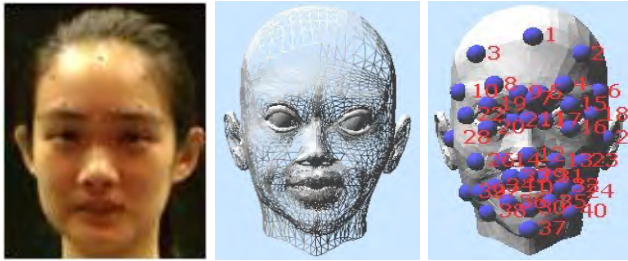
(a) 原始数据点云 (b) 识别的标记点 (c) 标记结构

图2 运动捕捉设备中重建的人脸表情模型

表情模型,从左到右分别为原始数据的点云,识别的标记点,以及带有数据清理的标记结构。

3.2 特征点的选取

特征点需要在能够表现人脸表情的关键特征区域选取,这些关键区域包括前额、左眉、左上眼睑、右眉、右上眼睑、左下眼睑、左脸、鼻子、右下眼睑、右脸、上嘴唇、下巴和下嘴唇。本文实验前在原始人脸模型上选取41个特征点,特征点的选取结果如图3所示。



(a) 正面照片 (b) 网格模型 (c) 特征点位置
图3 带有标记的正面照片、网格模型及41个特征点位置

3.3 人脸表情的克隆

人脸表情克隆主要目的是把源模型上已经存在的顶点运动向量等数据重新定位到目标模型上,并同时保持原始面部动画的相对运动、动力学特征。

对于每一种关键表情,通过计算源模型特征点的位移来参数化^[11]目标模型特征点的位移向量,然后根据目标模型41个特征点的位移,利用基于凸权值的拉普拉斯变形算法求出所有顶点变形后的3D坐标,从而得到目标模型的克隆表情。

4 实验结果与讨论

本文实验采用的计算机配置为1.83 GHz Intel Core Duo T2400 CPU,2 GB内存,捕捉一位女性表演者的六种基本表情序列作为源表情,选择一位外国男性gene和一只猴子作为目标模型,分别采用本文方法和文献[13]方法进行实验,同时选择一位顶点数目较多的中国男性模型与文献[5]方法进行对比实验,实验效果如图4所示。图中(a)为源模型的六种基本表情,从左到右依次为生气、厌恶、恐惧、高兴、悲伤、吃惊。(b)、(c)分别为采用本文方法和文献[13]方法在男性gene模型上得到的克隆表情。从(b)、(c)各列对应的克隆表情可以看出,特别是从第一列生气表情、第四列高兴表情和第六列吃惊表情的嘴部表情可以明显看出,文献[13]得到的克隆效果不够自然,有效性低。图(d)、(e)分别为采用本文方法和文献[13]方法在猴子模型上得到的克隆表情。观察(e)中各列对应的克隆表情,特别是第一列生气表

情、第四列高兴表情、第五列厌恶表情和第六列吃惊表情的逼真度低于(d)中各列对应的克隆表情,(d)中的实验结果表明了本文的方法是有效的,很好保留了人脸表情的细节,克服了文献[13]方法逼真度低的问题。图(f)、(g)分别为采用本文方法和文献[5]方法在中国男性模型上得到的表情克隆效果。本实验的中国男性模型含有8 171个顶点,15 974个三角形网格;外国男性gene模型含有988个顶点,1 954个三角形网格;猴子模型含有1 227个顶点,2 344个三角形网格。中国男性模型的顶点和网格三角个数远远超过了外国男性gene和猴子模型的个数,与图4中的源模型表情相比,(g)中第一、二、三列的眼部表情和第四、五、六列的嘴部表情在真实



(a) 源模型的六种基本表情



(b) 男性gene对应的六种克隆表情(本文算法)



(c) 男性gene对应的六种克隆表情(文献[13]算法)



(d) 猴子对应的六种克隆表情(本文算法)



(e) 猴子对应的六种克隆表情(文献[13]算法)



(f) 中国男性对应的六种克隆表情(本文算法)

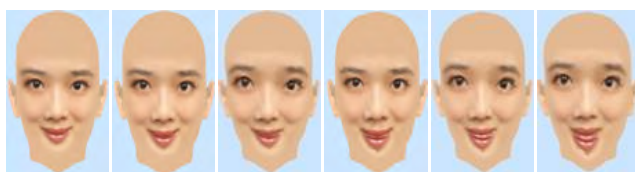


(g) 中国男性对应的六种克隆表情(文献[5]算法)

图4 六种表情的源模型和克隆后的目标模型

性上低于(f)各列对应的克隆表情,文献[5]方法得到的克隆表情不够真实、自然。正如文献[5]所指出的,当目标模型顶点数目较多,并且表情数据较稀疏时,该方法的克隆结果将会丢失源模型的部分表情细节。

为进一步形象直观地说明本文方法的有效性,以gene和猴子作为目标模型为例,选取吃惊表情具有代表性的六个关键帧进行实验,并与基于余切权值的拉普拉斯算法进行对比,实验效果如图5所示。图中(a)为源模型吃惊表情关键帧,从左到右,嘴部和眼睛的张开幅度逐渐增大,吃惊表情也逐渐增强。(b)、(c)分别为采用本文方法和余切权值在男性gene模型上得到的吃惊表情关键帧。从左到右,(c)中眼睛和嘴部依次张开的幅度没有(b)中明显,与源模型相比,(b)的吃惊表情更具有表现力。(d)、(e)分别为采用本文方法和余切权值在猴子模型上得到的表情克隆效果。余切权值在计算时可能为负,并且当角度接近 180° 时计算值不稳定,致使重建后的表情有效性降低。观察(d)、(e)的克隆效果,特别是从猴子嘴部张开的幅度可以明显看出,与基于余切权值的拉普拉斯算法相比,本文方法得到的克隆效果更真实、有效。



(a)源模型吃惊表情关键帧



(b)男性gene对应的克隆表情关键帧(本文方法)



(c)男性gene对应的克隆表情关键帧(余切权值方法)



(d)猴子对应的克隆表情关键帧(本文方法)



(e)猴子对应的克隆表情关键帧(余切权值方法)

图5 吃惊表情六个关键帧的源模型和克隆后的目标模型

5 总结

本文提出了一种数据驱动的人脸表情克隆新方法,易于编辑实现,可以有效地合成具有真实感的克隆表情。通过采用凸权值的拉普拉斯变形算法,尽量多地保持了源模型的局部表面信息,能够被快速有效地求解稀疏线性系统,同时将误差扩散到整个网格,有效提高了重建的效果,满足了普通用户对表情克隆的需求。但是本文给出的方法存在一定的局限,没有将头发考虑在内,不能重建得到带有头发的完整目标模型,同时没有考虑带有头部运动或语音的目标模型。这些问题将会在今后的工作中进一步解决。

参考文献:

- [1] Pyun H, Kim Y, Chae W, et al. An example-based approach for facial expression cloning[C]//Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation. Aire-la-Ville: Eurographics Association Press, 2003:167-176.
- [2] Noh J Y, Neumann U. Facial animation by expression cloning[D]. Los Angeles: University of Southern California, 2002.
- [3] Liu Zicheng, Zhang Zhengyou, Jacobs C, et al. Rapid modeling of animated faces from video[J]. The Journal of Visualization and Computer Animation, 2001, 12(4):227-240.
- [4] Zhang Zhengyou, Liu Zicheng, Adler D, et al. Robust and rapid generation of animated faces from video images: a model-based modeling approach[J]. International Journal of Computer Vision, 2004, 58(2):93-119.
- [5] Noh J Y, Neumann U. Expression cloning[C]//Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 2001:277-288.
- [6] Zhang Mandun, Yao Jian, Ding Bin, et al. Fast individual face modeling and animation[C]//Proceedings of the Second Australasian Conference on Interactive Entertainment. Sydney: Creativity & Cognition Studios Press, 2005:235-239.
- [7] Wan Xianmei, Jin Xiaogang. Spacetime facial animation editing[J]. Journal of Computer-Aided Design & Computer Graphics, 2013, 25(8):1183-1189.
- [8] Deng Zhigang, Neumann U. Expressive speech animation synthesis with phoneme-level control[J]. Computer Graph Forum, 2008, 27(8):2096-2113.
- [9] Seol Y, Lewis J P, Seo J, et al. Spacetime expression cloning for blend shapes[J]. ACM Transactions on Graphics, 2012, 31(2):1-12.

参考文献:

- [1] Fukunaga K, Hostetler L. The estimation of the gradient of a density function, with applications in pattern recognition[J]. *IEEE Transactions on Information Theory*, 1975, 21(1):32-40.
- [2] Cheng Yizong. Mean shift, mode seeking, and clustering[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, 17(8):790-799.
- [3] Comaniciu D, Ramesh V, Meer P. Real-time tracking of non-rigid objects using mean shift[C]//*IEEE Conference on Computer Vision and Pattern Recognition, Proceedings*, 2000:142-149.
- [4] Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(5):564-576.
- [5] Jeyakar J, Babu R V, Ramakrishnan K R. Robust object tracking with background-weighted local kernels[J]. *Computer Vision and Image Understanding*, 2008, 112(3):296-309.
- [6] 杜凯, 巨永锋, 靳引利, 等. 自适应分块颜色直方图的 Mean-Shift 跟踪算法[J]. *武汉理工大学学报*, 2012, 34(6):140-144.
- [7] Gouiffes M, Laguzet F, Lacassagne L. Projection-histograms for mean-shift tracking[C]//*IEEE 17th International Conference on Image Processing*, 2010:4617-4620.
- [8] Leichter I, Lindenbaum M, Rivlin E. Mean shift tracking with multiple reference color histograms[J]. *Computer Vision and Image Understanding*, 2010, 114(3):400-408.
- [9] Leichter I. Mean shift trackers with cross-bin metrics[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(4):695-706.
- [10] Fashing M, Tomasi C. Mean shift is a bound optimization[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(3):471-474.
- [11] Li X, Zhang T, Shen X, et al. Object tracking using an adaptive Kalman filter combined with mean shift[J]. *Optical Engineering*, 2010, 49(2).
- [12] Shen C, Brooks M, van den Hengel A. Fast global kernel density mode seeking: applications to localization and tracking[J]. *IEEE Transactions on Image Processing*, 2007, 16(5):1457-1469.
- [13] Li S X, Chang H X, Zhu C F. Adaptive pyramid mean shift for global real-time visual tracking[J]. *Image and Vision Computing*, 2010, 28(3):424-437.
- [14] Li S X, Wu O, Zhu C F, et al. Visual object tracking using spatial context information and global tracking skills[J]. *Computer Vision and Image Understanding*, 2013.
- [15] 沈云琴, 陈秋红. 自适应均值漂移算法目标跟踪检测仿真研究[J]. *计算机仿真*, 2012(4):290-292.
- (上接 181 页)
- [10] Bhat K S, Goldenthal R, Ye Yuting, et al. High fidelity facial animation capture and retargeting with contours[C]//*Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. New York: ACM Press, 2013:7-14.
- [11] Wan Xianmei, Jin Xiaogang. Data-driven facial expression synthesis via laplacian deformation[J]. *Multimedia Tools and Applications*, 2012, 58(1):109-123.
- [12] Sorkine O, Cohen-Or D, Lipman Y, et al. Laplacian surface editing[C]//*Proceedings of 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*. New York: ACM Press, 2004:175-184.
- [13] Yue Zhen. The research about individual facial molding[D]. Hangzhou: Zhejiang University, 2008.
- [14] Meyer M, Desbrun M, Schroder P, et al. Discrete differential-geometry operators for triangulated 2-manifolds[M]//Hege H C, Polthier K. *Visualization and Mathematics III*. Berlin: Springer Press, 2003:35-57.
- [15] Zhang Yongmian, Ji Qiang, Zhu Zhiwei, et al. Dynamic facial expression analysis and synthesis with MPEG-4 facial animation parameters[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2008, 18(10):1383-1396.